

METHOD AND DISPLAY FOR MULTIVARIATE CLASSIFICATION

Related Applications

This application claims priority under 35 U.S.C. §119 from U.S. provisional application serial number 60/240,403, filed October 13, 2000, and from U.S. provisional application serial number 60/252,290, filed November 21, 2000.

Field of the Invention

The invention relates to methods of classifying tissues based on multivariate data for diagnostic, prognostic or therapeutic purposes. The invention also relates to methods and user interfaces for displaying such multivariate data to facilitate rapid decision making.

Background of the Invention

Classification of tissues historically depended on an examination of the gross morphology and histology of the tissues. The inadequacy of these methods for classification of tissues that are similar in appearance, such as different tumors arising from the same tissue or organ, is known. More recently, methods of tissue analysis and classification have been developed that rely on genetic analysis of tissues. While these genetic analytical methods are more powerful for distinguishing tissue types, and are simple to practice, the methods also generated quantities of data that are orders of magnitude greater than classical histology methods. For example, gene expression profiling of tissues by the application of microarray technology determines simultaneously the expression of thousands of genes. The challenge of identifying the data useful in tissue classification from the raw data is a problem that requires a solution that can be applied in clinical settings. Present statistical and data display methods are not satisfactory for this application.

Thus it is of key importance in analytical processes for classification of unknown tissues reduce large quantities of multivariate data to a quantity that can be readily analyzed. A typical treatment of multivariate data is to generate binary statistics known to those in the art. For example, a binary regression of two data sets associated with two objects will produce a statistical relationship between the two objects within a statistical confidence.

Prior art data analysis tools have focused on describing objects of diagnostic relevance to exhaustion, which in turn relied on using binary statistical methods to "collapse" large volumes of data for decision making. These methodologies suffer from a requirement

for sophisticated analysis and interpretation by the end user. The prior art approaches thus lack methods for analyzing large quantities of multivariate data using statistically proven methods, yet providing for ready interpretation of the data.

Highly parallel quantitative measurement systems are increasingly available for analysis of complex biological systems, but practical data reduction systems limit application. For example, microarray (RNA) expression analysis is used for phenotyping of human tissues; quantitative measurement of thousands of RNAs ("variables") in a single tissue ("object") is used to assign group membership ("classify"). Typically, a very large number of variables are measured in a training set of two different tissues, and a subset of variables which best distinguishes the two tissue types is identified through some statistical process. This can yield dozens, or even hundreds, of variables which are individually imperfect, but collectively effective, for classification of future (test) objects which were not part of the training set.

Once a predictive model is constructed from a suitable training set, it is desirable to have a simple method to generalize this model to a variety of variable measurement systems (different manufacturer's gene arrays, for example). This is not readily possible if the different platforms have independent units of measurement, and different sensitivity thresholds. Diagnostic and prognostic evaluation of tissues is hampered by this cross-platform incompatibility. Similar problems exist with respect to sample-to-sample, run-to-run and user-to-user variability when using the same device or an identical device.

In addition, the variables are not easily evaluated for any sample, even with the training set data as a comparison. For example, gene expression data frequently is provided as a table describing the increases in expression of the set of genes. This type of display requires the end user to compare large sets of data for the expression of multiple genes and make judgments as to the phenotype of a tissue based on these multiple parameters. Accordingly, there is a need to develop methods of obtaining and analyzing data that describe a tissue more accurately and more precisely for diagnostic evaluation. Further, there is a need for methods of displaying multivariate data without reference to the specific units of data generated by specific analytical devices. There also is a need for methods and devices to display the data for rapid and simplified evaluation by end users, particularly in a clinical setting.

Summary of the Invention

The present invention represents a new approach to data analysis for multivariate classification, particularly as used in medical diagnostics. The invention is in part an intuitive decision making tool for rapid classification of "objects" (e.g., cell, tissue or tumor samples) from evaluation of many simultaneous "variables" (e.g., quantitative gene expression profiles). The data analysis methods of the invention provide the end user with a simplified and robust output for diagnostic classification of objects based on identifying and evaluating multiple variables of predetermined diagnostic relevance. The raw data generated by analysis of the variables is transformed by application or appropriate algorithms to scaleless rank differentials between the variables. The rank orders of variables are used to classify tissues based on readily observable user interfaces, such as a graphical (e.g., visual) user interface or an auditory user interface.

According to one aspect of the invention, methods of classifying an object are provided. The methods include the steps of measuring, among one or more samples of the object, values of a plurality of variables associated with the object; selecting one or more of the plurality of variables by which the object may be identified; and representing the object by the one or more variables.

In certain embodiments, the step of selecting includes selecting the one or more variables on the basis of a difference between values of the variables and values of variables of another object. In other embodiments, the step of selecting further comprises steps of comparing the values of the plurality of variables with values of a plurality of variables associated with another object and selecting the one or more of the variables by which the object may be identified in response to the comparing step. Preferably the object is a tissue, and the step of measuring comprises microarray expression phenotyping the tissue.

In additional embodiments, the step of selecting comprises calculating mean values of variables among variables of two groups; comparing the mean values; and selecting the one or more of the variables by which the object may be identified in response to the comparing step. The step of selecting also can include ordering the variables according to the calculated mean values. Preferably the step of ordering comprises normalizing values of the variables to a predetermined range.

According to another aspect of the invention, additional methods of classifying an object are provided. The methods include steps of calculating a rank order associated with index set values (TRANK); calculating a second rank order associated with the difference of

index and contrasting set values (DIFFTRAN); calculating index set means (ALLTMEAN); and comparing these rank orders and means to previously indexed objects. In certain embodiments the methods include displaying, to a user, a visual indication of a relationship described by the rank orders. Preferably, the visual indication is a ternary plot.

5 According to still another aspect of the invention, methods of characterizing data sets containing multiple variables are provided. The methods include statistically determining a subset of the variables to describe a difference between the data sets; in a first set of data, ranking each of the variables in the subset of variables according to its numerical value in the first set of data; calculating a difference for each of the variables in the subset of variables between the first set of data and a second set of data; and ranking each of the variables in the subset of the variables according to the difference for each of the variables in the subset of variables. Preferably the methods include displaying the subset of variables of the first set of data according the rankings of each variable of the subset.

10 Methods for rendering, to a user, a computer interface are provided in other aspects of the invention. The methods include obtaining a set of variables associated with an object; reducing the set of variables to a subset of variables, which are associated with the object by a predetermined parameter; ranking each of the variables in the subset of variables according to their numerical value; and applying a computer interface associated with a human sense. In some embodiments, the human sense is hearing, vision, and touching.

15 In yet other aspects of the invention, methods for classifying an object are provided. The methods include obtaining a set of variables associated with an object; ranking each of the variables according to their numerical value; and providing an human interface through which a user may interact.

20 In still other aspects of the invention, methods of classifying an unknown tissue are provided. The methods include measuring values of each variable of a set of variables for an index group of tissues and a contrast group of tissues, calculating mean value and differences of mean value for each variable of the set of variables from the index group of tissues and the contrast group of tissues, ranking the means and differences of the means between the index and the contrast groups, determining values of each variable in the set of variables in an unknown tissue, and comparing the measured values in the unknown tissue to rank orders selected from the group of ranked means of the index group, ranked means of the contrast group, and differences of the means between the index and the contrast groups.

09977054-101201

In some embodiments, the methods also include displaying to a user a visual indication of a relationship described by the comparison of the measured values of variables in the unknown tissue and the rank orders of the index and contrast groups. Preferably the methods include coding display indicators, such as symbols or tones, based on the positive or negative values of the means and the differences of the means. In certain embodiments, the visual indication is a ternary plot. In preferred embodiments, the variables are genes and the values are levels of gene expression; more preferably, the values are expressed as log of gene expression.

According to still other aspects of the invention, computer-readable media are provided that are encoded with one or more programs that, when executed on a computer system, perform one or more of the foregoing methods. Computer systems for implementing such methods also are provided in accordance with the invention.

These and other aspects of the invention will be understood with reference to the drawings and the detailed description below.

Brief Description of the Drawings

Fig. 1 depicts ternary display of the ALL vs. AML model defined by ALL(n=27) vs. AML(n=11) teaching sets. The model description is based upon plotting of Index tissue rank order (ALLTRANK) and Rank order of difference between Index and Contrasting Tissue (DIFFTRAN). Two plots are shown, one for each tumor type. When the ALL teaching set means (ALLTMEAN) are plotted on the third axis (left panel), a linear vertical array of datapoints results. When the AML teaching set means (AMLTMEAN) are plotted (right panel) on the third axis, the cloud of circles migrates to the left and cloud of crosses to the right. Notice also the number of symbols on the axes. Circles = genes expressed higher in AML; Crosses = genes expressed higher in ALL.

Fig. 2 depicts a challenge of the ALL vs. AML model with new test cases. Ternary plots of the model defined in Fig. 1 (right and bottom axes) were plotted against 20 new ALL (top) and 14 new AML (bottom) tissues. Note in the ALL tissues most circle symbols are in the vertical midline or on the right axis. In the AML tissues, the circle symbols tend to move away from the right axis, and veer to the left of midline. In the ALL tissues very few of the cross symbols are on the right axis - most are in the midline. In the AML tissues these move to the right and begin collecting on the right axis.

Fig. 3 depicts ternary display of two breast cancer types. 100 genes which distinguish two types of breast cancer (designated H and L) were used to develop a model (upper row). Individual tissues of known classification which were part of the training set are shown below. Note data cloud shifts in individual tissues reflect trends summarized in the model.

Figs 4-6 depict plots of the same data analyzed in Figs. 1-3, using an alternative rank order plot. In these plots, the rank order of the contrasting tissue expression is plotted on the Y (right) axis. As in Figs. 1-3, the rank order of the index tissue expression is plotted on the X (bottom) axis, and the log transformed data of unknown (challenge) tissue gene expression is plotted on the Z (left) axis.

Fig. 7 shows a listing of the 100 genes which best distinguish AML and ALL, as selected by standard statistical approaches.

Fig. 8 depicts a listing of the 100 genes which best distinguish AML and ALL, and their rank orders in AML and ALL, along with expression values for each AML and ALL test tissue in additional columns.

Detailed Description of the Invention

The invention described herein relates to the classification of objects of medical diagnostic relevance. The object can be a tissue, fluid or other bodily component, such as normal or cancerous tissues. The invention will be described primarily as it relates to tissues, although it will be understood that the invention is more broadly applicable to other bodily components and objects.

In certain embodiments of the invention, an initial evaluation is generated using a training set of objects. As used herein, a "training set" or "teaching set" is a group of objects, such as tissue samples, of known kind (e.g., phenotype). Preferably the tissues used for generating training sets are classified using standard methods such as histology and/or antibody recognition. The training set of objects is assayed for desired variables to establish patterns of data that serve to distinguish the training set of objects from a different kind of objects. In some embodiments, two representative objects of two distinct classes of objects are identified, although it generally will be preferred to assay a larger number of representative objects to establish variables characteristic for each type of object. For example, when examining tissue samples, multiple samples of known histological type are assayed for a plurality of variables.

Preferably, a large number of variables associated with the two classes of objects are measured, and a subset of these variables which best distinguishes the two objects is identified through one or more statistical processes. Collectively these variables describe the distinction between the two objects, preferably within a predetermined statistical confidence range. The subset of the variables can be used subsequently to classify unknown objects presumed to be a member of one of the two distinct classes.

The data acquired for the variables are transformed to a scaleless rank order. This transformation eliminates units of measurement, thus providing cross-platform comparability. For example, data generated for gene expression analysis of tissues by different manufacturer's fluorescent microarray readers can be directly compared following the data transformation, preferably following calibration using reference gene analysis. Thus training sets generated by a certain type of microarray reader can serve as comparative data for diagnosis of unknown tissues using expression data generated using a different type of microarray reader.

Examples of variables include: gene expression (e.g., data obtained via DNA microarray or protein microarray), tissue histology morphometry (nuclear size, chromatin texture index, percentage distribution of cell types), tissue or cell growth conditions and properties (*in vitro* culture nutritional requirements, cell division rate, etc.).

The robustness and utility of the model are attributable to several unique features including the following. First, incorporation of many variables permits accurate classification of objects based upon predominant trends between variables rather than reliance on one or a small number of variables. This "fuzzy" classification protocol is well suited to overlapping states (nonexclusive variable states for different objects) and rapid sensory (e.g., visual or auditory), rather than algorithmic, interpretation of data. Second, the ternary plotting system is refractory to linear scale changes. Once the model is constructed, the data entered from a test object is displayed in an identical fashion irrespective of multiplicative scaling of the variables. Third, the model used to define the display is extremely simple, based solely on rank ordering of quantitative index variables in two prototypical object states. It will be possible to translate the model between measurement platforms (from an Affymetrix expression system, for example, to a glass spotted microarray) by simple recalibration of variable rank orders in a series of calibrators. These calibrators could be a series of representative samples, or pooled specimens, measured in compared systems. Fourth, although the display described herein is configured for bimodal

classification of objects with only two data states (e.g. normal tissue vs. tumor tissue, or tumor having good prognosis vs. tumor having bad prognosis), it could be expanded by performing a series of pairwise classifications, or increasing dimensionality (e.g., a three dimensional pyramid plot). Fifth, the output patterns are similar for a variety of models, thus achieving a degree of visual standardization for classification decisions irrespective of the tissues being studied. Sixth, incomplete test tissue data (e.g., missing values for some genes) has very little impact on classification performance when applied against a defined predictive model. Because the classification decision is based on displacement of a datapoint cloud, the reduced cloud density conferred by missing datapoints has no effect on the positions (and thus shifting) of the cloud itself.

The display of data clouds in a ternary (or other) plot permits the use of image recognition software to aid in classification. For example, a standard pattern (e.g., image) recognition software package can be linked to the data display to recognize or aid in the recognition of the tissue type being classified. The pattern recognition software can compare the data or image generated against training set data or images to differentiate between tissue types, or to suggest to the user a possible classification. The pattern recognition software can also be linked to a database containing standardized training set data or images of a variety of tissue types to facilitate the recognition of unknown tissue types. For example, if a tumor biopsy is isolated from a subject that does not present typical gross morphology or histology, a clinician may not be able to provide a suggestion as to classification of the tumor. In such cases, where the tumor biopsy is then screened by a methodology generating multivariate data, such as microarray gene expression profiling, the clinician or laboratory technician may not be able to select an appropriate group of training sets against which to compare the image or data generated by the method of the invention. In such cases, pattern recognition software linked to a database may be able to search for a data or image pattern from a database containing many training set patterns, thereby providing the clinician or laboratory technician with a classification or a small group of potential classifications from which the classification can be made.

In certain embodiments, software for performing the statistical manipulations described herein (transformation to scaleless data) can be provided on a computer connected by data link to a data generating device, such as a microarray reader. Any standard data link can be used, including serial or parallel cables, radio frequency or infrared telemetry links. Alternatively, data can be transferred by computer-readable medium (e.g., magnetic or

optical medium) and read by the software. The data also can be entered directly by the user via user interface, such as a keyboard, monitor, etc. The computer may be contained within the data generating device, providing an integrated system for generating raw data, transforming the data to scaleless data, and displaying such data. One or more computers also may be linked to one or more data generating devices and one or more display devices, such as in a local area network or wide area network.

After acquiring the raw data from the data generating device, the data for the variables examined can be transformed to scaleless data (e.g., rank order data) in accordance with the methods of the invention. The software can allow the user to select a number of variables preferred for classification, or the software may provide scaleless data for a standardized set of variables (e.g., genes known to be useful for classification of a tissue type or set of tissue types). The software can execute data transformation algorithms from a preselected group, or can allow the user to input other algorithms. The transformed scaleless data can be stored in a data file, printed, and/or directly displayed to the user on a graphical user interface.

In one embodiment of the invention, a visual display is used to display the scaleless data for the classification of tissues. A ternary plot can be used as the visual display, provided to the user by a graphical user interface, such as a monitor, or a printer.

Other user interfaces known to one of ordinary skill in the art also can be used. For example, in other embodiments, an auditory display can be provided to the user to display the scaleless data for the classification of tissues. For these embodiments, the data for the variables is converted to distinct tones or series of tones. Thus a series of tones can be assigned according to an assignment scheme, with variable (e.g., a gene) representing a tone of a given pitch, to provide an acoustic scenario which may be easily recognized.

In one embodiment, the assignment of the series of tones is in rank order of expression within the training set index tissue and the series of tones is a progressive scale sequence. A different series of tones is assigned to the independently rank ordered contrasting tissue set, to yield a distinctly different melody or sequence. An unknown tissue then is assigned a series of tones based on the rank order of variable expressions in the unknown tissue. In another embodiment, the series of tones for the index tissue and/or contrasting tissue gene expression data is a non-scalar melodic pattern.

Other modifications to the tones may be made to differentiate the data of different sets. For example, certain auditory qualities can be modified, such as producing tones of different duration and/or intensity based upon a measure of proximity of rank order between

the unknown tissue and the index tissue. Similarly, duration and/or intensity of tones can be varied based on the proximity of rank order between the unknown tissue and the contrasting training tissue.

The user can compare the tones generated by the software for the tissue being classified with training set tissues to determine the classification of the tissue. The user can select a best match between the unknown tissue and one of the acoustic signatures (tonal patterns) for the index and contrasting tissue classes. Other displays also may be used, including multimedia displays combining visual and auditory outputs described above.

Examples

As a proof of concept, gene expression in tissues of Acute Lymphocytic Leukemia (ALL) and Acute Myelogenous Leukemia (AML) were classified. The gene expression data show that the two tumor types have systematic differences in gene expression. The data published by Golub et al. (*Science* 286:531-537, 1999) were used; expression of 5000+ genes was analyzed from 27 (ALL) and 11 (AML) training tissues.

The 100 genes which best distinguish the two groups were selected by standard statistical approaches (Fig. 7). First, the mean expression values of the two groups were compared, and only those with a minimum-3-fold and 100-unit difference between means were retained. The logged means were then rank ordered by t-test and the top 50 genes which were upregulated in AML, and the top 50 upregulated in ALL were selected for modeling. This represented 100 genes, or variables, which distinguish the two groups of AML and ALL. The objective was to develop a model of how these variables can distinguish the object types, or tissue groups, and apply that model to classify new, or test, tissues.

The model was built using the mean values for each gene in the two tissues, and is completely described by only two variables: 1) the rank order of gene expression in the Index Tissue (TRANK), and 2) the rank order of the difference in expression (DIFFTRAN) between the two tissues (Index and Contrasting Tissue).

First, one of the two tissues was selected to be the Index tissue. The 100 variables in the index tissue were rank ordered by the mean expression of each to generate an Index Rank. The ALL training set was used to calculate an Index Rank for 100 genes in ALL. This was simply the rank order of genes sorted by their average expression amongst the 27 training ALL tissues. Rank ordering stripped away all absolute scale information.

Second, the Contrasting Tissue (AML) was compared to the Index Tissue (ALL), again using the mean expression for each gene. The genes rank were ordered by the difference between the means (Contrasting Tissue mean - Index Tissue mean; or AML - ALL). This supplied a scaleless rank-order distance metric between the genes called the

5 DIFFTRAN.

Ternary Display of the Model:

The next step in the classification analysis was to use the model, composed entirely of two values (TRANK and DIFFTRAN) for each of 100 genes, to classify a test tissue for which the 100 gene measurements were available.

10

Ternary, or triangular, coordinate displays were used to plot the two-dimensional model (dimensions 1 and 2, TRANK and DIFFTRAN) against the test, or unknown tissue (dimension 3). Ternary displays permit plotting of three variables, each on a separate axis, in only two dimensions. This is possible when the three variables are known to have a fixed relationship, such as when they always add to the same number.

15

Ternary Plots are suited to binary object classification when the model is defined by variables with a fixed relationship to each other. In general, the following relationships hold true in the model:

$$\text{Test Case} = \text{TRANK} + \text{DIFFTRAN}.$$

20

If the test case was a member of the index group, then $\text{DIFFTRAN} = 0$, and $\text{Test Case} = \text{TRANK}$. If the test case was a member of the Contrasting Tissue Group then:

$$\text{Test Case expression} - \text{DIFFTRAN} = \text{TRANK}$$

25

Thus, in cases where DIFFTRAN is not equal to zero (Fig. 1, Right), the plot axis rotates perpendicular to the axis when $\text{DIFFTRAN}=0$ (Fig. 1, Left).

Using the Model for Prediction: Ternary Plot of Unknown Tissue Against Classification

30 Model

The ALL vs AML Model defined in Fig. 7 was used to generate a model template against which new test tissues could be displayed. A simple spreadsheet in which 100 rows each represent one gene define the model in two columns containing ALLTRANK and

DIFFTRAN. Symbol color for each of the 100 genes was defined by the training set as red or blue, based upon direction of change between groups ("red" indicating genes expressed higher in AML; "blue" indicating genes expressed higher in ALL; these were graphed as circles and crosses, respectively, to show the differentiation in black and white figures).

Expression values for each test tissue appear in additional columns (Fig. 8). Fig. 2 shows the data cloud patterns in ternary plots as predicted for ALL (top) and AML (bottom) test tissues, and should be compared to the training set shown in Fig. 1. ALLTRAN (bottom axis) and DIFFTRAN (right axis) rank order values for each of 100 genes were calculated from a training set of 27+11 tissues, and plotted against the test tissue expression value (Left Axis) for each of the 100 genes. These plots were generated using an actual dataset, in which Fig. 2 test tissues had no input into defining the model (training sets).

The axes for Fig. 2 shows AML and ALL test cases. Bottom Axis: TRANK = Rank order of expression level in Index Tissue. Right Axis: DIFFTRAN = Rank order of difference between Index and Contrasting Tissue. Left Axis: Log of expression in Unknown Tissue. Symbol shape was used to show direction of difference in model, although other variations of the data cloud (e.g., color of data points) would work equally well.

When the AML teaching set means (AMLTMEAN) are plotted (Right) on the third axis, the circle symbol cloud migrates to the left and cross symbol cloud to the right. Visibility of the axis associated datapoints (symbols on the axes) can be improved by slight jittering, to reduce overlap. The differences in the data clouds in these ternary plots permit ready differentiation of ALL from AML tumors by just glancing at the plots.

Fig. 3 shows comparable ternary displays for an entirely different tumor type, breast cancer, emphasizing that the display appearance is maintained independent of the genes selected, or tumor type studied. The individual cases shown in Fig. 3 (all but the top row) are the actual training cases used to generate the model shown in the top row. In this case, the individual patterns reproduce the trends seen in the average model.

Another embodiment of the two-class model is based solely on the rank order of gene expression within the two compared training set groups. In this approach, a list of discriminating genes was selected as above from a training set using standard statistical approaches. Gene expression for each tissue class was rank ordered as above by abundance, and these two model elements (rank order of index tissue group, RANK-I; rank order of contrasting tissue group, RANK-C) were plotted against an unknown sample (log transformed). Figs. 4-6 are examples plotted in this manner using a triangular, or ternary,

coordinate system from the same datasets shown in Figs. 1-3, respectively. The X (lower) axis shows rank ordered training set index tissue, Y (right) axis shows rank ordered training set contrasting tissue, and Z (left) axis shows unknown or challenge tissue. In particular, this model displays excellent graphic displacement of datapoints representing genes (cross symbols) expressed at a higher level in index (X axis) compared to contrasting (Y axis) training set tissues. Differences in these data presentations may be observed by comparing the shift of cross-symbol data clouds between Figures 1 and 4, 2 and 5, 3 and 6.

Data Processing Flowchart

1. Model defined by gene (a,b,c,d,...) expression in Training Set of 2 tissue classes:

Index and Contrasting tissues.

Calculate expression mean and differences for expression of each gene a,b,c,d...

$$I_{a,b,c..} = \text{Index group mean}$$

$$C_{a,b,c..} = \text{Contrast group mean}$$

$$C_{a,b,c..} - I_{a,b,c..} = D_{a,b,c..} = \text{Difference of means}$$

2. Convert 2-variable Model to (lowest-highest) rank order:

$$I_{r_{a,b,c..}} = I_{a,b,c..} \text{ rank order}$$

$$C_{r_{a,b,c..}} = C_{a,b,c..} \text{ rank order}$$

$$D_{r_{a,b,c..}} = D_{a,b,c..} \text{ rank order}$$

3. Symbols coded by positive/negative direction of Dm

$$D_m < 0, \text{ symbol} = \text{circle}$$

$$D_m > 0, \text{ symbol} = \text{cross}$$

4. Log transform unknown tissue to be plotted against model = U

Log Transform U;

5. Each gene (a,b,c,...) is plotted as one symbol, in three dimensions on a ternary (triangular) plot. Two dimensions define the model (X,Y), and one dimension (Z) is the challenge case.

Coordinate	Gene a	Gene b	Gene c
X (lower) axis	Ir_a	Ir_b	Ir_c
Y (right) axis	Dr_a	Dr_b	Dr_c
Z (left) axis:	$\log_2(U)_a$	$\log_2(U)_b$	$\log_2(U)_c$

6. Interpret Graphic Signature Created by Plotting Unknown against model

Examples shown in Figs. 1-3

Another Embodiment:

Step 5 of the flowchart was modified as follows. The Rank order of the Contrasting tissue expression ($Cr_{a,b,c,...} = Cm_{a,b,c,...}$ rank order) is plotted on the Y axis. Thus:

Coordinate	Gene a	Gene b	Gene c
X (lower) axis	Ir_a	Ir_b	Ir_c
Y (right) axis	Cr_a	Cr_b	Cr_c
Z (left) axis:	$\log_2(U)_a$	$\log_2(U)_b$	$\log_2(U)_c$

Interpretation according to step 6 of the flowchart yields the exemplary graphs shown in Figs. 4-6.

Each references cited herein is hereby incorporated by reference.

While the invention has been described with respect to certain embodiments, it should be appreciated that many modifications and changes may be made by those of ordinary skill in the art without departing from the spirit of the invention. It is intended that such modification, changes and equivalents fall within the scope of the following claims.

What is claimed is: